
GSA
U.S. General Services Administration
FedFleet
2020
Federal Automotive Statistical Tool (FAST)
Continuing the Focus on Data Quality
Ron Stewart
INL/CON-20-57196:001

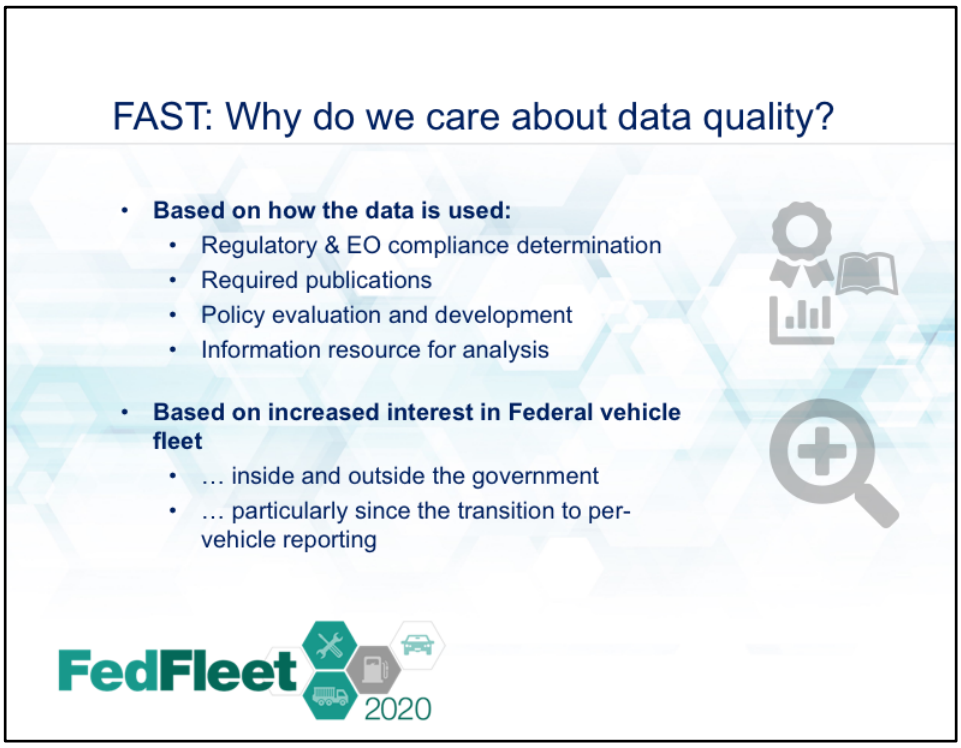
- Overview of FAST and how vehicle fleet data is reviewed
- Preliminary review of FY 2019 fleet dataset
  - How does it compare to prior year?
  - Where do we see trouble spots?
- Recommendations
- Discussion

FedFleet 2020

What is FAST?

- **Collaborative project funded and managed by**
  - GSA's Office of Government-wide Policy
  - DOE's Federal Energy Management Program
  - Energy Information Administration
- **Collects multiple federal vehicle fleet datasets**
  - Vehicles & operational data (costs, miles, fuel)
  - Fleet budget data
  - Fueling infrastructure data

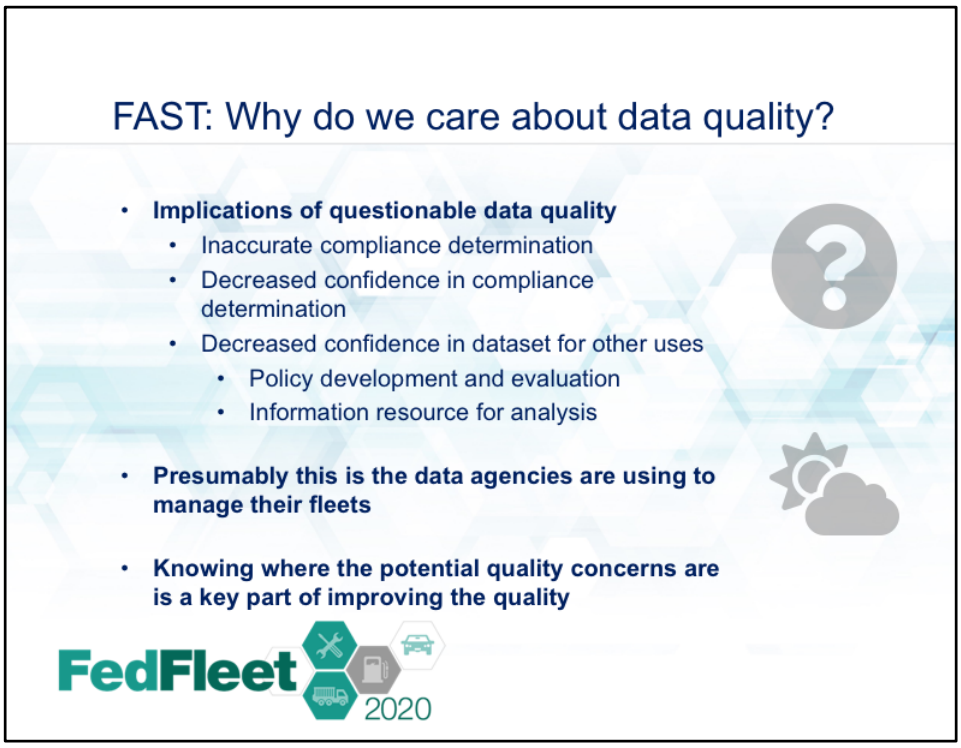
FedFleet 2020



Quality: within this context, we can think of "data quality" in terms of whether the data represents a complete and correct picture of the Federal fleet and its operation for the corresponding fiscal year.

Complete: all the vehicles that were part of the fleet, and all the needed information about each vehicle

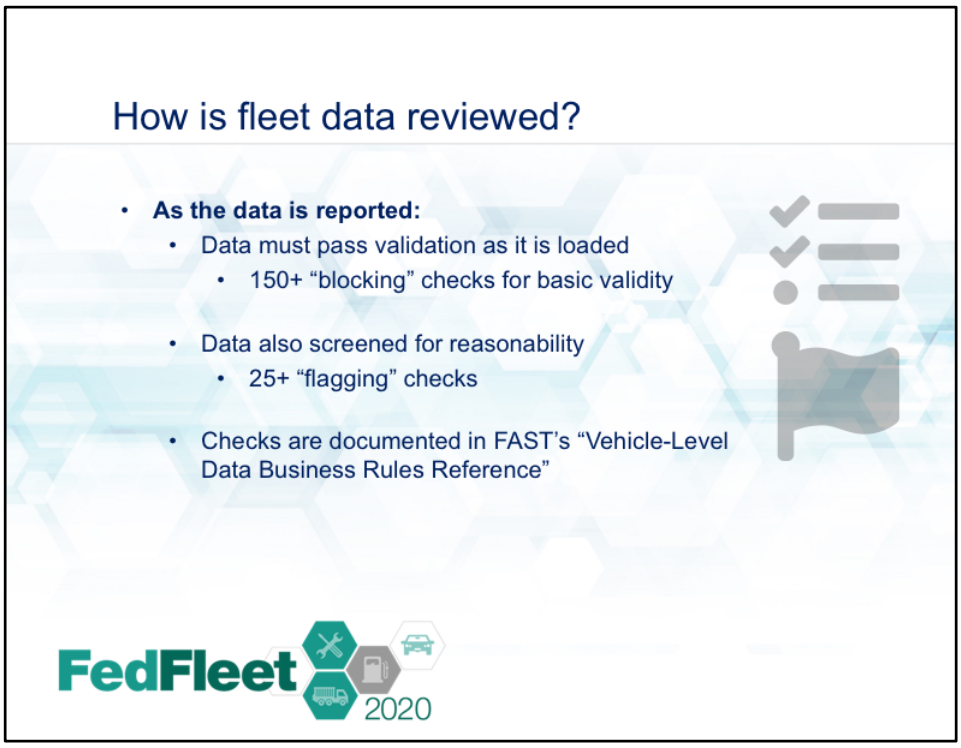
Correct: all the information provided for a given vehicle is accurate



Without reasonably complete and correct data about a given agency's fleet, it is quite possible that any determination of compliance with a given requirement is inaccurate.

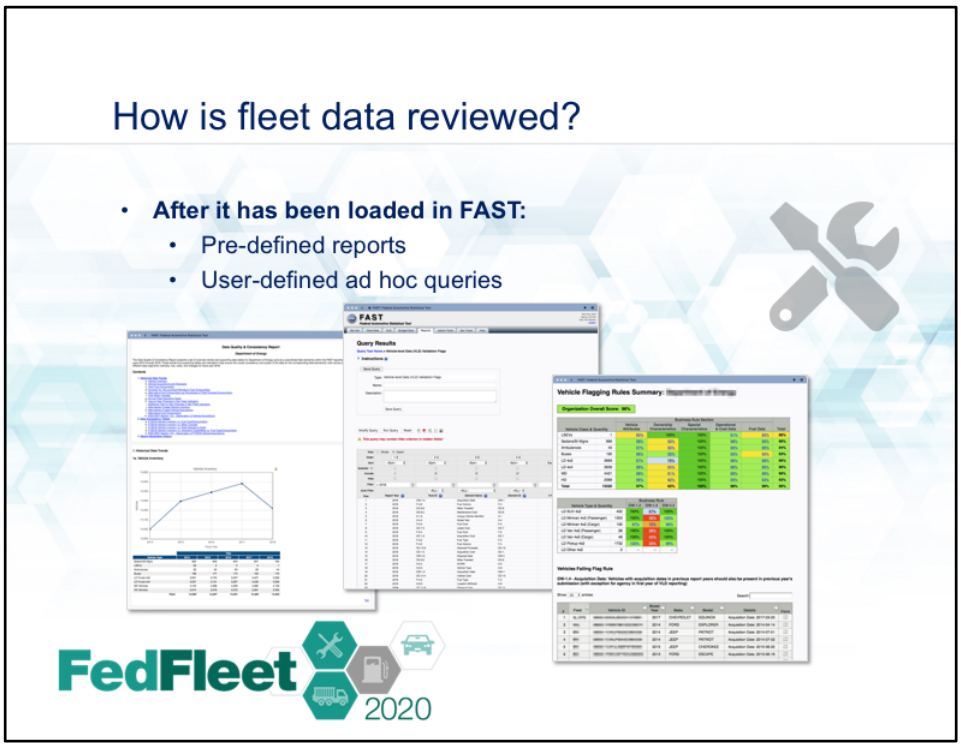
If there are questions about whether a given agency's submission is complete and correct, it lowers the confidence in any compliance determination (including for other agencies).

Questionable quality also makes it more difficult to defend or support use of the data for its other intended purposes



150+ validation checks for basic validity:
These look at aspects such as required information being provided, correct information types (numbers vs. text), values within required ranges or from specified sets, and consistency of groups of attributes

25+ validation checks for reasonability:
These look at whether the information falls within expected ranges for a "typical" federal fleet vehicle

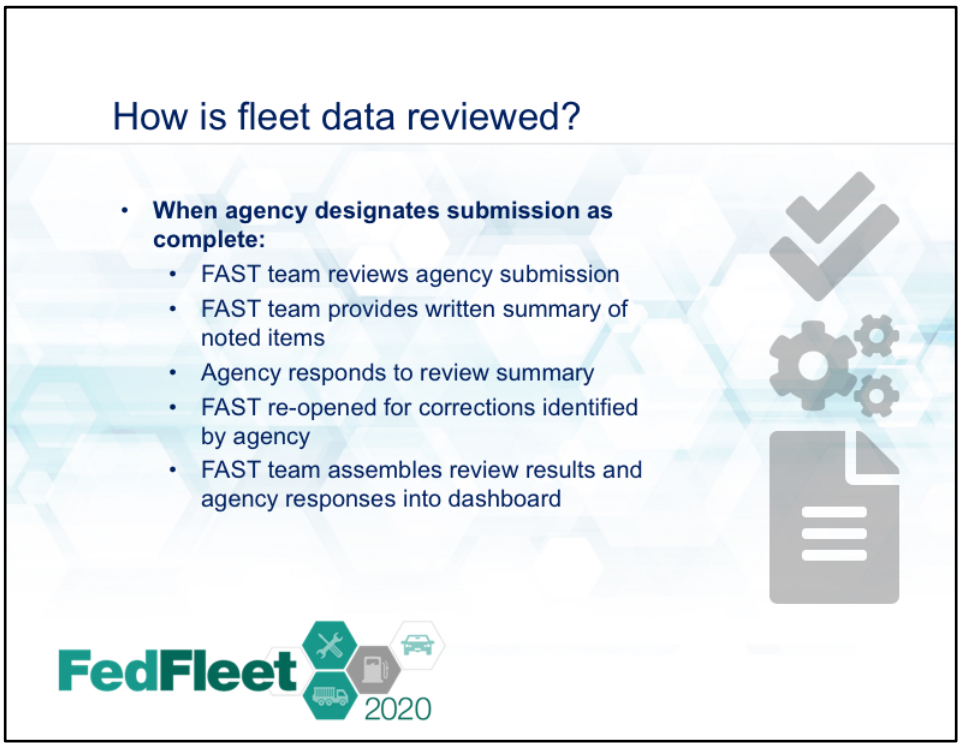


FAST has several reports that agencies can use to review the data after it has been loaded in FAST:

- Data Quality & Consistency Report (macro look at many of the same metrics we will talk about in a few minutes); can be used to look at agency or subordinate organizations; can also be used to look at specific subsets of the vehicles within those portions of the fleet (e.g., just the LE vehicles or just the foreign vehicles).
- Flagging Rules Summary: VLD-specific report enables digging into specific flagging rules for vehicles failing those flagging checks

FAST's query tool has several VLD-specific query types:

- Vehicles
- Fuel
- Flagging failures



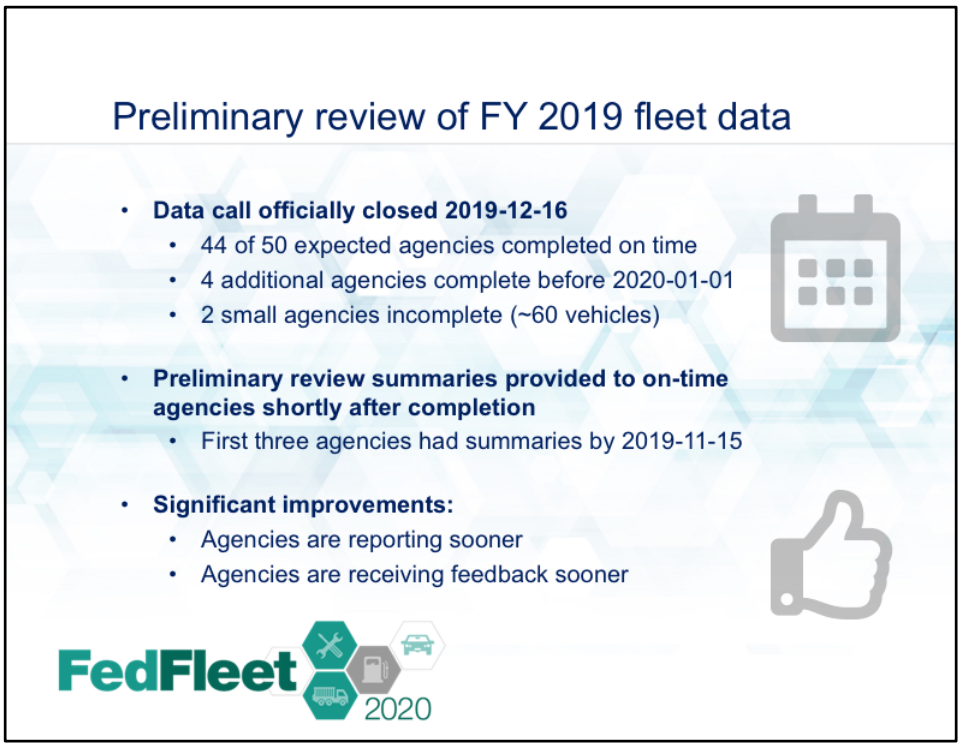
This is the same basic review/response process the FAST team initiated last year.

The dashboard – as we will see in a couple minutes – is very important from a long-term standpoint, so that we can have a reasonably clear picture of where quality issues are and whether or where data quality is changing over time (by comparing year to year).

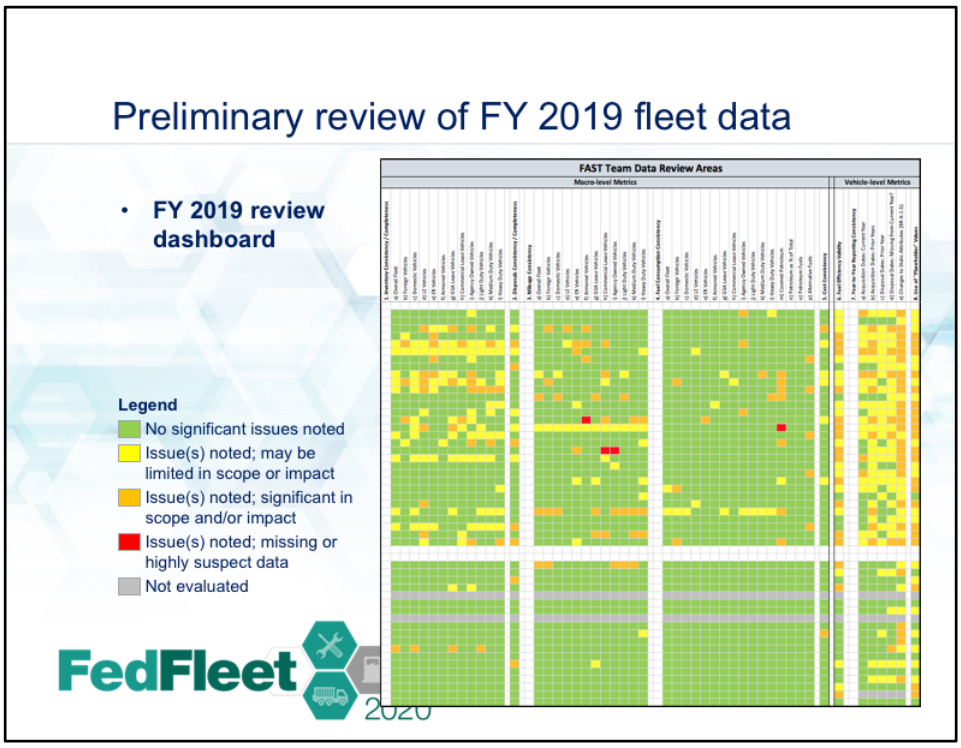
How is fleet data reviewed?

- **FAST team review looks at two levels**
  - Macro: high-level consistency
    - Unexpected shifts in fleet size or makeup?
    - Inconsistent shifts in costs, miles, fuel?
  - Micro: vehicle-level consistency and trouble spots
    - Year-to-year consistency
    - Fuel efficiency (e.g., mileage or fuel)
    - Use of “placeholders” for vehicle information
- **DOE review looks primarily at compliance**
  - Identifies issues impacting compliance

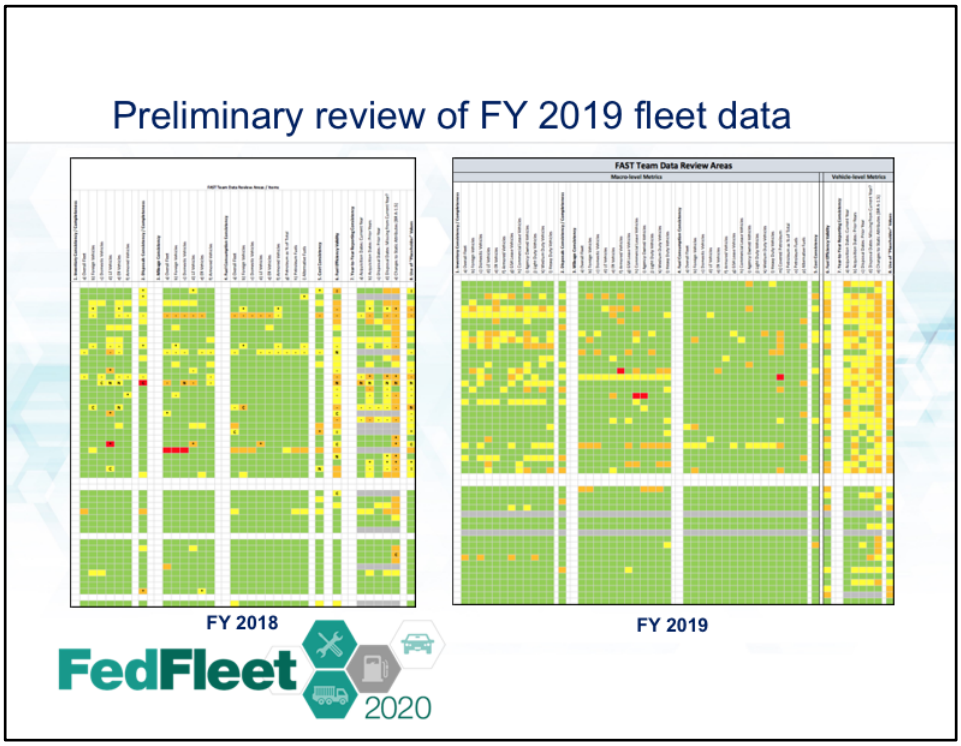
FedFleet 2020



Each of those agencies receiving early feedback (mid-November) was able to address some or all of the concerns noted in the preliminary review feedback



Everything discussed from this point forward is based as of 01/13/2020, when agencies began receiving review summaries and before any agencies began submitting either their responses to that review or corrections to their fleet data submissions. All of this is subject to change based on both of those (agency responses, agency corrections).



Discuss areas where we see changes from FY 2018 to 2019:

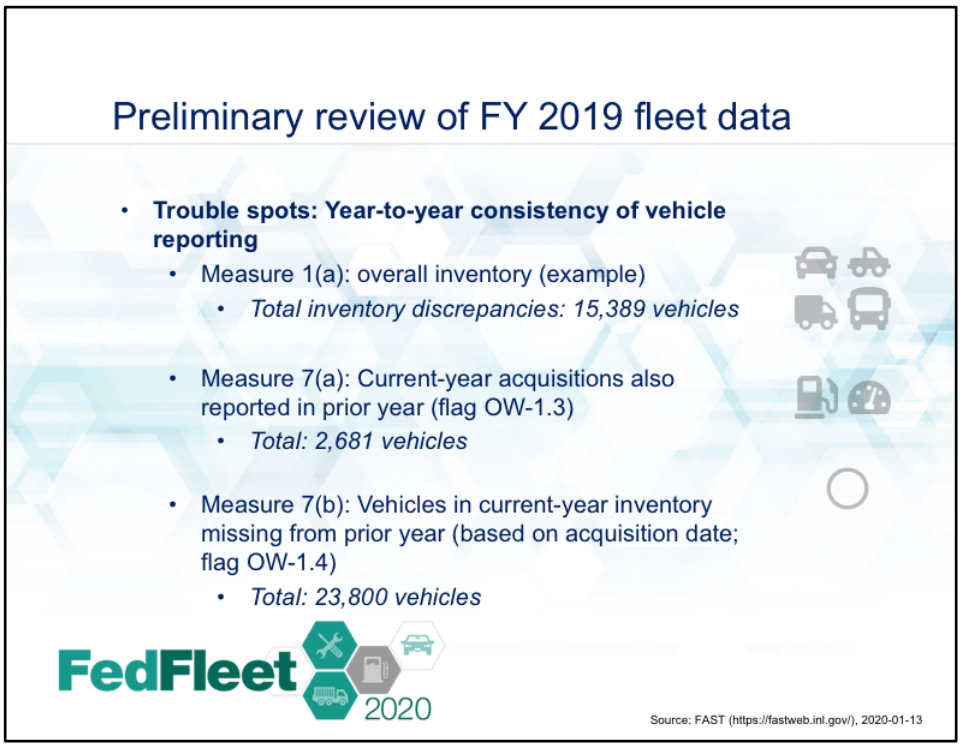
- Fewer major (red) items this year than last (good)
- More yellow and orange, although that is likely to change based on agency responses and corrections before we consider FY 2019 data and review final

Summary: Difficult to look at this comparison and feel like FY 2019 is better than 2018.

Preliminary review of FY 2019 fleet data

- **Where do we see trouble spots?**
  - **M1 & M7: Year-to-year consistency of vehicle reporting**
    - Missing / inconsistent reporting of vehicles
    - Changes in "static" vehicle attributes
  - **M6: Fuel and mileage at the vehicle level**
  - **M8: Use of "placeholder" values for vehicle attributes**
    - Particularly for vehicle fuel and mileage

FedFleet 2020



Measure 1(a): this looks at each agency's full fleet and how consistent last year's inventory, this year's acquisitions and disposals, and this year's inventory all are.

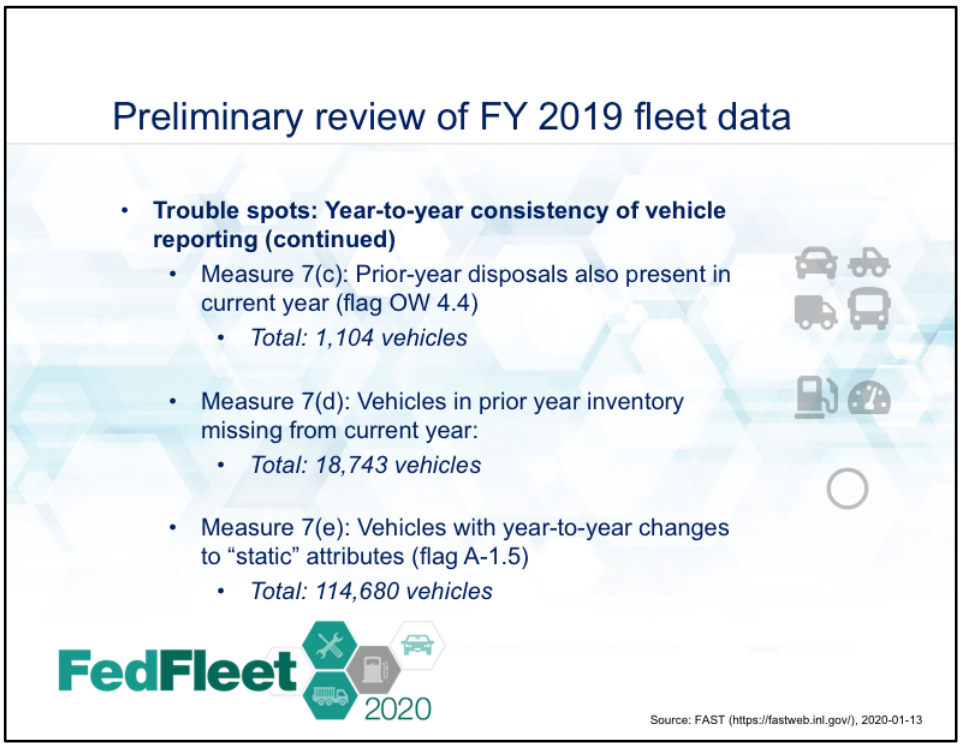
Example: if agency reported 1,000 vehicles in inventory last year, and the agency reports 100 disposals and 50 acquisitions this year, but they also report 1,100 vehicles in inventory this year, that's a discrepancy of 150 vehicles (1,000 PYI – 100 CYD + 50 CYA = 950 expected CYI vs 1,100 reported CYI)

Measure 1 has other subsets (foreign vs domestic, by ownership, by weight class, by category (LE, ER, armored) that it also looks at. We've used versions of this metric for at least the past 10 years as a macro-level look for issues. It makes it relatively easy to spot situations like an agency neglecting to report their foreign inventory or their owned inventory (both of which have happened as recently as last year). Most of the measure 1 metrics can also be found in some way on FAST's Data Quality and Consistency Report.

Measures 1 and 7 are very closely related: measure 1 is a macro look, where measure 7 looks at different aspects of consistency of vehicle reporting at the individual vehicle level, and any significant inconsistencies at the individual vehicle level will almost certainly contribute in some way to macro-level issues.

Measure 7 metrics rely heavily on year-to-year constancy of vehicle ID's, but we do take a "soft match" approach to trying to find vehicles within an agency's submission in the current year and prior year (e.g., VIN's with/without org prefix).

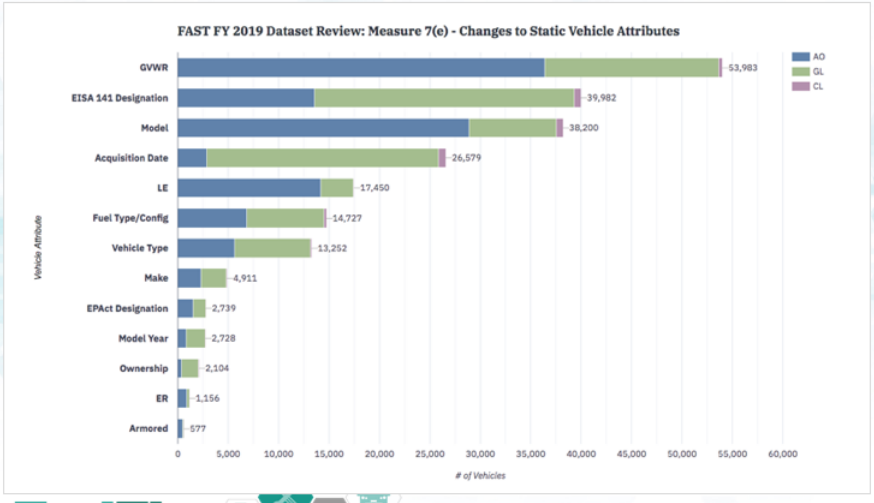
Each of the metrics within measure 7 look for specific situations that can cause problems (explained on this and next slide).



4 of 5 metrics within measure 7 correspond to flags which are generated during loading of the data; agencies have access to that flagging data during their reporting and pre-completion review – so there shouldn't be surprises there when agencies receive review summaries.

The exception is 7(d) which really can't be checked until an agency's submission is complete. (More on that on coming slide…)

Preliminary review of FY 2019 fleet data

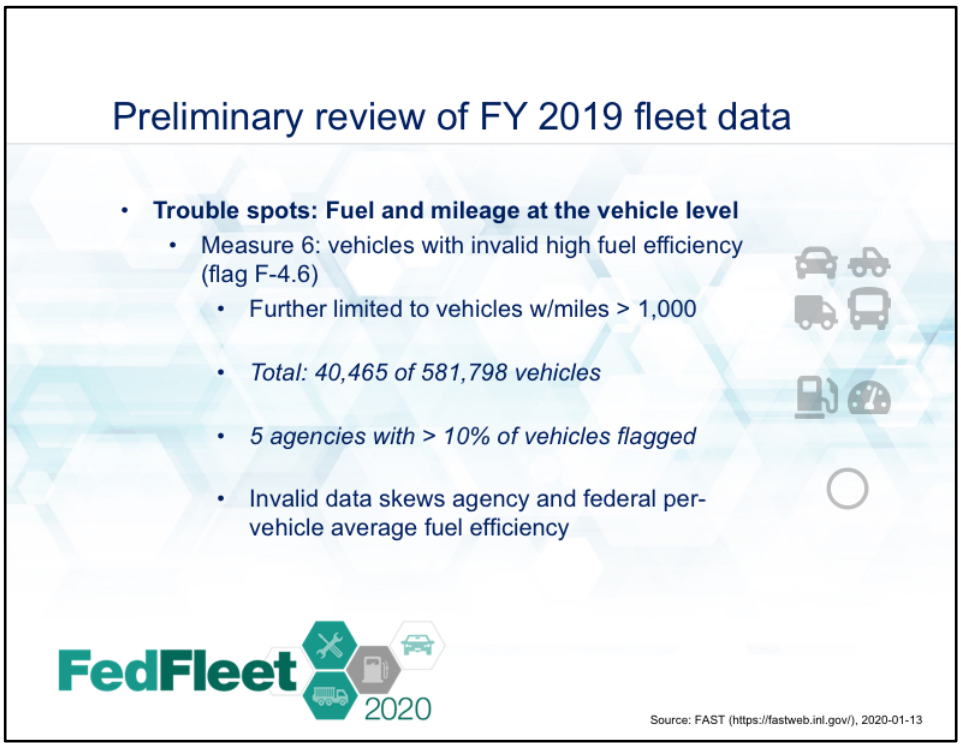


FedFleet 2020

Source: FAST (https://fastweb.inl.gov/), 2020-01-13

Looking a bit at some additional detail on what sorts of things are showing up in measure 7(e), some possible causes, and why some of them might – or might not be – of concern: this chart is a breakout of each of the different attributes this metric (and flag A-1.5) looks at, based on how many vehicles showed changes in that attribute, and broken down by ownership type.

Discuss which raise potential concerns, but also why having some of these change from year-to-year may be a good thing (e.g., if the changes are a result of getting more accurate information as agencies continue data clean-up… such as GVWR)



Flag F-4.6 establishes ranges by vehicle class for "reasonable" fuel efficiency; measure 6 takes a subset of those: only vehicles with 1,000 miles or more for the year (so that low-mileage vehicles don't inflate the numbers based on timing of fueling at the beginning and end of the year), and further limits it to just those flagged for being above the high threshold for reasonability. That flag also already has some special handling in place for EV's and plug-in hybrids to account for their potentially higher fuel efficiencies.

But we still ended up with over 40,000 vehicles identified with questionable high fuel efficiency… which can be traced back to issues with fuel consumption and/or mileage for each vehicle.

5 agencies (all large or very large) had more than 10% of their vehicles fall into this category.

Significant issues here can easily skew the average per-vehicle fuel efficiency for an agency, and there are enough of them that it clearly skews that average for the Federal government.

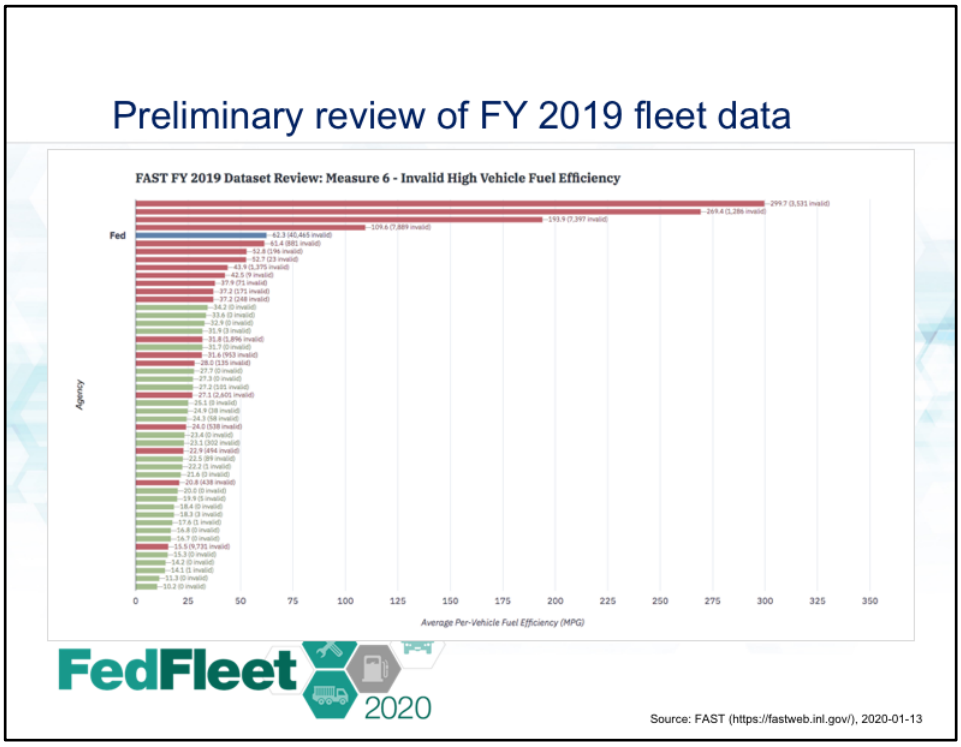
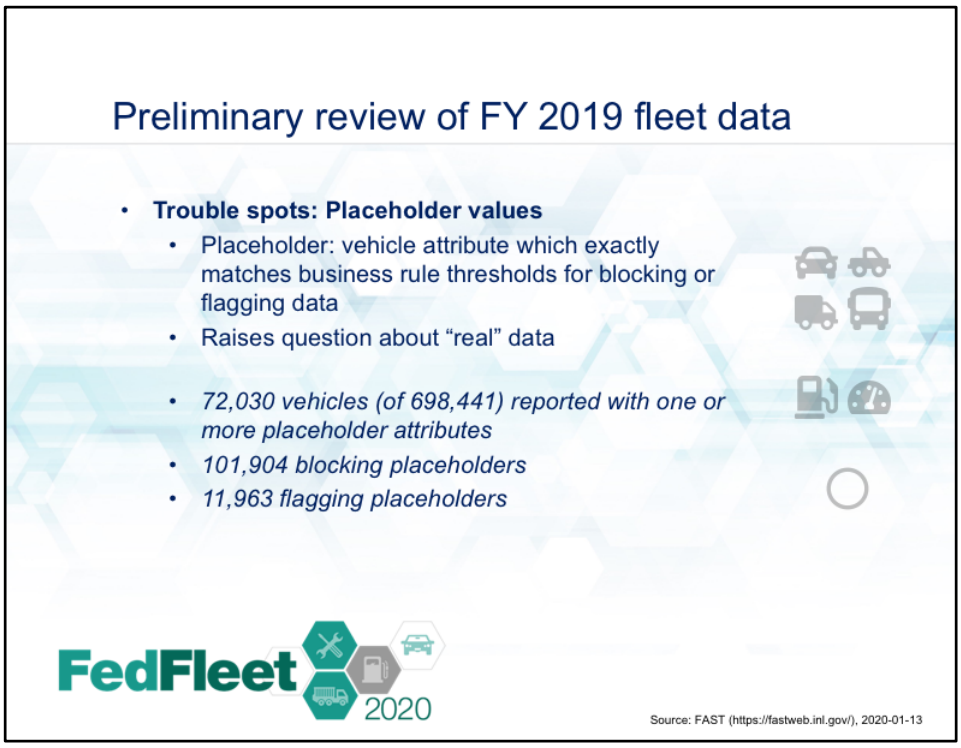


Chart shows – without identifying specific agencies – where all agencies fall

Blue line is the federal average across all vehicles @ just over 62 miles/gallon. That has to be considered suspect based on what we know about the makeup and operation of the fleet.

Red bars represent agencies with suspect averages based on
- their average (e.g., anything above ~35 miles/gallon for an agency per-vehicle average is questionable),
- or large numbers or percentages of vehicles with individually invalid fuel efficiencies.

Nearly half of the agencies show signs of this skewing, even some with lower averages (based on the number or percentage of vehicles identified by this metric).



Placeholder example: BR A-5.2 and A-5.3 enforce absolute limits of 500 and 100,000 lbs for vehicle GVWR. If we find a vehicle reported with one of those values, that would be considered a blocking placeholder. A-5.4 enforces a flagging threshold of 80,000 lbs; if we find a vehicle reported with a GVWR of that value, it would be considered a flagging placeholder.

It's important to note that the data may be right, but the fact that a vehicle attribute exactly matches the threshold raises questions about whether the information is "real" or was just selected by the reporting organization in order to get the data to load into FAST. There are some situations where we can have a fairly high degree of confidence that the placeholder is in fact that (e.g., a vehicle reported with a vehicle ID that is based on its VIN, and decoding the VIN yields information about the year, make, model, etc., that is consistent with those attributes, but the GVWR is reported as 500 lbs), or if there are significant numbers of vehicles reported with identical values.

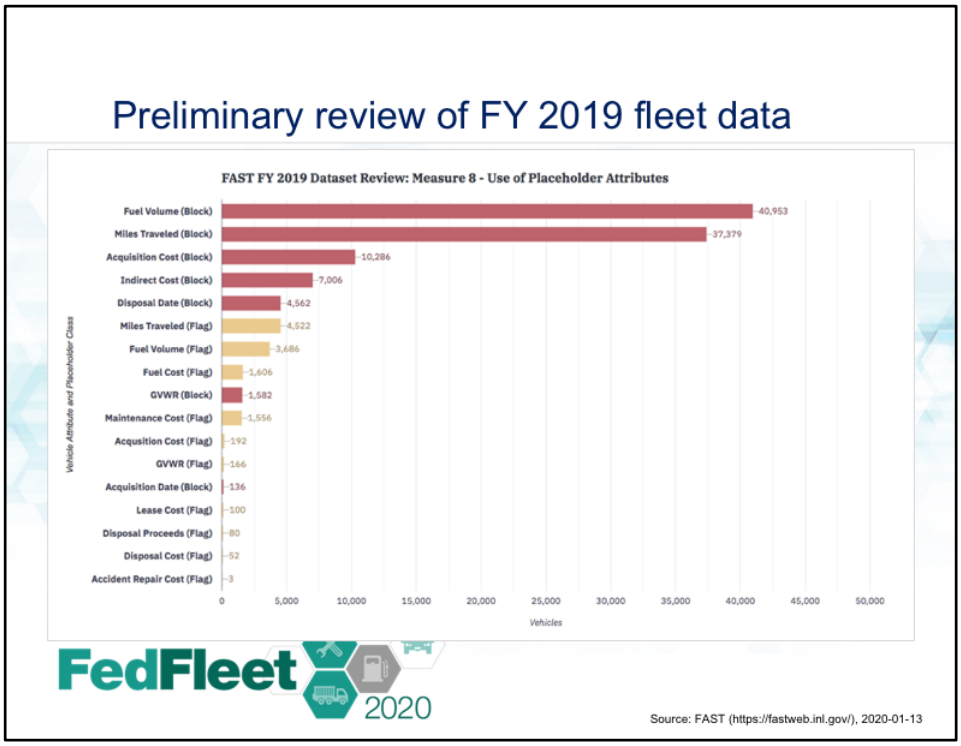
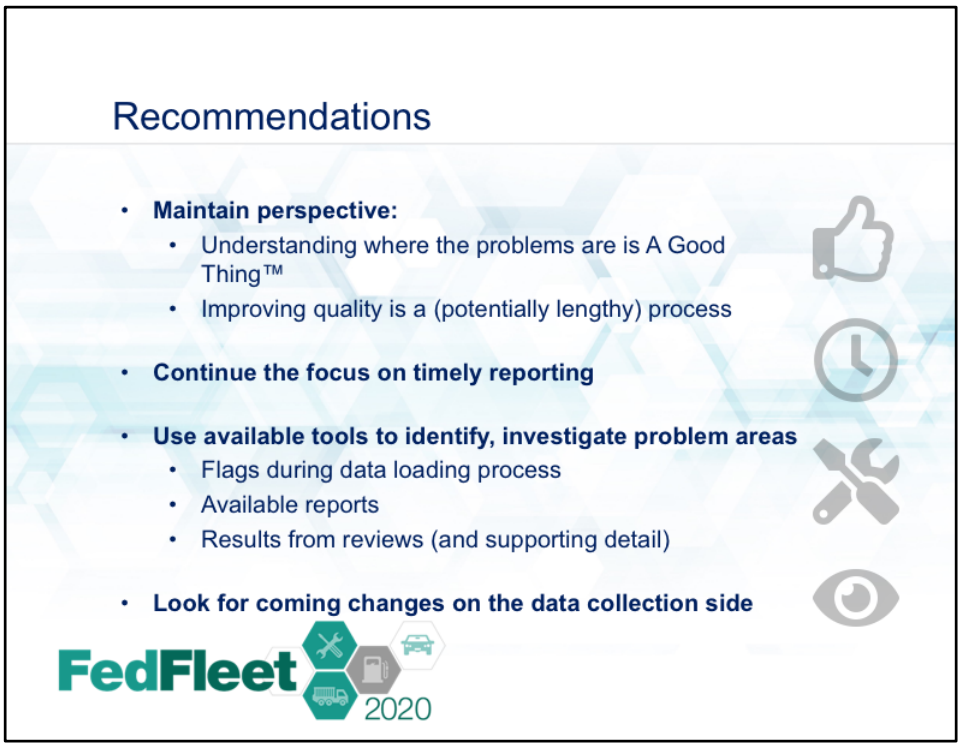


Chart shows the vehicle attributes where our review found what appears to be use of placeholder values, and the number of vehicles with placeholders on each attribute.

Blocking placeholders shown in red; flagging placeholders shown in yellow.

Particularly striking here are the top two: vehicles reported with blocking placeholders on fuel volume (almost all reported as 1 unit of fuel) and miles traveled (again, almost all reported as the minimum allowed). And there are smaller groups with flagging placeholders on these same attributes.

This is the primary contributor to the number of vehicles with invalid fuel high fuel efficiencies.



Importance of perspective:

- Easy to see those last few slides touching on the problems that we see and think the sky is falling and everything is horrible.
- There are clearly problems, but…
- The reality is that many of the problems we're seeing now because we have vehicle-level data have been there all along… but now we know they are present and we have a better sense of where they are, and we can begin working to address them.
- We also recognize that for a fleet of any size, improving the accuracy and completeness of the fleet data is potentially a lengthy process… but the result is important: better data to manage the fleet with (in addition to more accurate data to report).

Timely reporting: if we look at when agencies are completely their initial submission over the past three years, the improvement is impressive. As the reporting processes the agencies use get smoother, they involve less time on the agencies part. It also indicates that the agencies are arriving at a point where their data is at least valid enough to load into FAST much quicker (that by itself is indicative of an improvement in quality), and it means we can provide feedback earlier within the reporting period.

Use the available tools: many of the key things that the FAST team's review touches on are available to you on the agency side as you go through the reporting process (existing reports, information in the flags that are generated, the query tool). - -

- Consider vetting your data against other related datasets (e.g., fuel consumption data reported to FAST vs tools like DOE's FleetDASH).
- Use the review summaries to decide which are the key items (or even the easier items, the low-hanging fruit) to address in the coming year.
- Use prior year review summaries to give you an idea of where past trouble spots were.
- Use the FAST sandbox as early as possible to ensure basic validity and to get an early perspective on where there may be groups of vehicles being flagged to particularly key issues like fuel and mileage.

Coming changes: there are several coming initiatives in the next year or two that will help in different ways

- The DOE and FAST team have talked about a significant overhaul of the EPAct 701 waiver request process that should significantly lighten the reporting burden on the agencies while also providing a better characterization of compliance. That could free up resources that can be focused on other data quality improvements within the agency.
- I touched on a couple of these metrics we look at in our review that are currently not available to agency users during the reporting process; we will be looking at incorporating those into tools available to FAST users to help with the agency reviews before they designate their submissions as complete.

- We're also looking at a couple of other changes in how we use the data which may ultimately result in simplifying reporting as well as more accurately characterizing agency compliance with some of the underlying requirements (that's a teaser for the February INTERFUEL meeting where we will spend a little bit of time talking about ideas in that area)


Discussion
FedFleet
2020

FAST Program Contacts

GSA OGP
- James Vogelsinger
  james.vogelsinger@gsa.gov
- Patrick McConnell
  patrick.mcconnell@gsa.gov

DOE FEMP
- Kendall Kam
  kendall.kam@ee.doe.gov
- Jay Wrobell
  jay.wrobell@ee.doe.gov

EIA CES
- Cynthia Sirk
  cynthia.sirk@eia.gov

FAST System Team
- Ron Stewart
  ron.stewart@inl.gov
- Michelle Kirby
  michelle.kirby@inl.gov
- Tim Raczek
  timothy.raczek@inl.gov
- @FASTdevs on Twitter

FedFleet 2020